

RELACIONAMENTO DE MÉTODOS DE AGRUPAMENTO POR CORRELAÇÃO DE MATRIZES DE VALORES COFENÉTICOS

Rubens Alves da Cunha ¹

ABSTRACT

RELATIONSHIP OF CLUSTERING METHODS BY CORRELATION OF COPENETIC MATRICES. Clustering methods are used in order to evaluate their performances using the same data set. The variables are used in their original scale or transformed by standardization and the similarity between pairs of OTU's is estimated by correlation and taxonomic distance. The resulting similarity matrices are clustered by different methods. The resulting matrices of copenetic values are correlated and these values rearranged in matricial form. These matrices are then clustered by UPGMA. The results showed the cases of better agreement between pairs of clustering methods according to the scale of the variables and to the measure of similarity between pairs of OTU's.

KEYWORDS. Copenetic correlation, Clustering methods, Phenetics.

INTRODUÇÃO

Quando, para uma dada situação experimental uma análise de agrupamento é indicada, torna-se necessário um procedimento que leve em conta a possibilidade de escalonamento de variáveis, a conseqüente seleção do coeficiente de semelhança e a adequação do método de agrupamento. Nesta situação, os dados experimentais são dispostos em forma matricial do tipo $D(n,t)$ onde n , número de linhas, corresponde a atributos e t , número de colunas, corresponde às OTU's (Unidades Taxonômicas Operacionais), ou seja, elementos que se pretende agrupar. Ocorre que, freqüentemente, os resultados a partir de $D(n,t)$ não são totalmente concordantes. Justifica-se pois um estudo sobre os resultados proporcionados por diferentes métodos de agrupamento, considerando-se ainda o escalonamento das n variáveis e a medida da semelhança entre OTU's.

Neste projeto, a avaliação dos resultados de agrupamento é feita com base em correlação das matrizes de valores cofenéticos resultantes da aplicação de diferentes métodos. A conceituação de correlação cofenética é devida a SOKAL & ROHLF (1962), como medida da representatividade da matriz de semelhanças entre OTU's pelo agrupamento correspondente. As técnicas de agrupamento, em sua grande maioria, estão descritas em SNEATH & SOKAL (1973), ANDERBERG (1973) e DUNN & EVERITT (1982).

¹. Departamento de Estatística Aplicada e Computacional / IGCE - UNESP, Rio Claro. SP, CEP 13500-230, Brasil.

MATERIAL E MÉTODOS

Serviram de suporte ao desenvolvimento deste projeto dados morfométricos de *Bombus atratus* Franklin 1913, utilizados por KERR & CUNHA (1990) para a obtenção de estruturas classificatórias, através de análises de agrupamento e de ordenação. Este conjunto de dados é representado por $n=43$ variáveis quantitativas (caracteres), relacionadas em KERR (1987) e $t=72$ unidades taxonômicas (OTU's), das quais 23 exemplares são machos, 27 são fêmeas e 22 são rainhas da espécie.

Para a análise pretendida, as variáveis são utilizadas em sua escala original ou transformadas por estandardização. A partir destas matrizes são calculados os coeficientes de correlação linear e de distância taxonômica entre OTU's. Estas matrizes de semelhanças (72×72) são então agrupadas por ligação simples (SINGL), ligação completa (COMPL), pelo método ponderado (WPGMA) e não ponderado (UPGMA) de agrupamento aos pares por médias aritméticas e pelo método ponderado de agrupamento com o recurso das somas de Spearman (WPGMS), quando a semelhança entre OTU's é medida por correlação linear. Para semelhança entre OTU's medida por distância taxonômica, são utilizados os métodos SINGL, COMPL, WPGMA, UPGMA e os métodos ponderado (WPGMC) e não ponderado (UPGMC) de agrupamento pela média de centróides. Em cada caso, uma matriz de valores cofenéticos é obtida. Em seguida, todas as matrizes de valores cofenéticos são correlacionadas e os resultados reunidos em forma matricial. Estas matrizes são então submetidas a agrupamento por UPGMA.

RESULTADOS

Do procedimento adotado, resultaram duas matrizes de correlação entre OTU's, sendo uma obtida a partir de variáveis em escala original e outra com as variáveis transformadas por estandardização. Estas matrizes foram submetidas a cinco métodos de agrupamento, quando a semelhança entre OTU's foi medida por correlação e a seis métodos, quando foram estimadas distâncias taxonômicas entre OTU's. De cada aplicação de um dos métodos de agrupamento resultou uma matriz de valores cofenéticos.

As matrizes de valores cofenéticos foram então correlacionadas, segundo o coeficiente de semelhança utilizado; os resultados obtidos dispostos em forma matricial são os constantes das tabelas I e II, nas quais um R ou um S acrescidos às siglas dos métodos de agrupamento indicam, respectivamente, variáveis em escala original e estandardizadas. Na seqüência, estas matrizes de correlação (tab. I, II) foram agrupadas por UPGMA e os fenogramas ilustrativos das inter-relações dos diferentes métodos de agrupamento são os da figuras 1 e 2. Em 1 o coeficiente de correlação cofenética é $r = 0,90$ e em 2 $r = 0,67$.

Tabela I. Matriz de correlação entre matrizes de valores cofenéticos dos diferentes métodos de agrupamento. Correlação entre todos os exemplares.

	SINGLR	COMPLR	UPGMAR	WPGMAR	WPGMSR	SINGLS	COMPLS	UPGMAS	WPGMAS	WPGMSS
SINGLR	1,000									
COMPLR	0,977	1,000								
UPGMAR	0,783	0,888	1,000							
WPGMAR	0,984	0,999	0,875	1,000						
WPGMSR	0,973	0,998	0,899	0,998	1,000					
SINGLS	0,445	0,568	0,821	0,550	0,579	1,000				
COMPLS	0,497	0,593	0,743	0,579	0,598	0,811	1,000			
UPGMAS	0,479	0,602	0,841	0,584	0,612	0,962	0,874	1,000		
WPGMAS	0,502	0,619	0,835	0,602	0,628	0,942	0,895	0,994	1,000	
WPGMSS	0,492	0,617	0,865	0,599	0,629	0,947	0,830	0,977	0,963	1,000

DISCUSSÃO

É relativamente comum que, ao serem apresentados os resultados de uma análise de agrupamento, a ilustração seja feita através de fenogramas, aos quais são associados os coeficientes de correlação cofenética. Este coeficiente é interpretado como nível de representatividade da matriz de semelhanças, pois é calculado entre os elementos desta matriz e os elementos da matriz resultante da aplicação de um método de agrupamento (matrizes de valores cofenéticos).

Resultados de uma análise de agrupamento podem, e freqüentemente ocorre, levar a diferentes grupos de OTU's, dependendo da escala da variável (caráter); do número deles que, se pequeno proporciona configurações instáveis, da medida de semelhança entre OTU's e do método de agrupamento utilizado.

No caso presente, uma comparação específica entre cada par de métodos de agrupamento pode ser obtida consultando-se uma das tabelas, respeitadas as condições em que foram obtidas. Por outro lado, os fenogramas ilustram, além dos casos de maior concordância entre métodos, a estrutura de todo o conjunto. Nota-se que na fig.1 estão representados dois grandes grupos de métodos, com semelhança próxima de $r = 0,60$, reunindo, de um lado, resultados a partir de variáveis em escala original e, de outro, a partir de variáveis estandardizadas. Internamente, os grupos de métodos se assemelham em torno de $r = 0,85$. Na fig. 2 tais agrupamentos não se revelam. Ainda com referência às figs.1 e 2, pelo menos dois outros aspectos merecem comentários. Quanto ao nível geral de semelhanças, o fenograma em 1 mostra maior homogeneidade do conjunto de métodos, o que pode também ser observado pelos resultados reunidos na tabela II. Ainda,



Figs. 1-2. Fenogramas (UPGMA) ilustrativos das inter-relações de métodos de agrupamento por correlação de matrizes de valores cofenéticos. 1) Correlação entre OTU's. Correlação cofenética $r = 0,90$. 2) Distância taxonômica entre OTU's. Correlação cofenética $r = 0,67$.

Tabela II. Matriz de correlação entre matrizes de valores cofenéticos dos diferentes métodos de agrupamento. Distância taxonômica entre todos os exemplares.

	SINGLR	COMPLR	UPGMAR	WPGMAR	UPGMCR	WPGMCR	SINGLS	COMPLS	UPGMAS	WPGMAS	UPGMCS	WPGMCS
SINGLR	1.000											
COMPLR	0.977	1.000										
UPGMAR	0.994	0.994	1.000									
WPGMAR	0.990	0.979	0.988	1.000								
UPGMCR	0.994	0.993	0.999	0.987	1.000							
WPGMCR	0.978	0.933	0.959	0.966	0.959	1.000						
SINGLS	0.997	0.964	0.985	0.986	0.985	0.980	1.000					
COMPLS	0.980	0.968	0.979	0.982	0.979	0.955	0.978	1.000				
UPGMAS	0.998	0.978	0.993	0.990	0.993	0.975	0.998	0.983	1.000			
WPGMAS	0.987	0.973	0.984	0.989	0.984	0.964	0.986	0.988	0.990	1.000		
UPGMCS	0.998	0.975	0.992	0.989	0.992	0.977	0.998	0.984	0.999	0.989	1.000	
WPGMCS	0.988	0.939	0.967	0.978	0.967	0.985	0.993	0.968	0.987	0.976	0.988	1.000

o coeficiente de correlação cofenética menor em 2, $r = 0,67$, pode ser visto como consequência de operação entre grandezas próximas, que não permitem encontrar grupos bem definidos, um dos objetivos da análise de agrupamento.

Assim, se resultados de métodos de agrupamento devam ser comparados, mostram-se mais semelhantes COMPL e WPGMA para correlação entre OTU's, com as variáveis em escala original, e UPGMA WPGMA para correlações entre OTU's, com as variáveis transformadas por standardização (fig. 1). No mesmo sentido, mas utilizando-se de distância taxonômica entre OTU's, os resultados dos mais concordantes são obtidos entre UPGMA e UPGMC, independentemente da escala das variáveis, o que ocorre também com os resultados da utilização do WPGMA (fig. 2).

Agradecimentos: À Universidade Estadual Paulista - UNESP, Rio Claro - SP (proc. 187/71 -IV) e ao "Centro di Ecologia Teorica ed Applicata", Gorizia, Itália, pelas condições de realização de estágio de pós-doutorado. Ao CNPq, pelo apoio financeiro (procs. 200 586-81.5 e 300 213/90-5)

REFERÊNCIAS BIBLIOGRÁFICAS

- ANDERBERG, M. R. 1973. **Cluster Analysis for Applications**. New York, Acad., 359 p.
- DUNN, G. & EVERITT, B. S. 1982. **An Introduction to Mathematical Taxonomy**. London, Cambridge Univ. 152 p.
- KERR, W. E. 1987. Sex determination in bees XVII. Systems of caste determination in the Apinae, Meliponinae and Bombinae and their Phylogenetical implications. **Revta bras.Genet.** Ribeirão Preto, **10**:685-694.
- KERR, W. E. & CUNHA, R. A.. 1990. Sex determination in bees XXVI. Masculinism of workers in the Apidae. **Revta bras. Genet.**, Ribeirão Preto, **13**:479-489
- SNEATH, P. H. A. & SOKAL, R. R. 1973. **Numerical Taxonomy**. San Francisco, W.H. Freeman, 573 p.
- SOKAL, R. R. & ROHLF, F. J. 1962. The comparison of dendrograms by objective methods. **Taxon**, Washington- DC, **11**:33-40.

Recebido em 04.05.1995; aceito em 13.02.1996.